

DOCUMENT RESUME

ED 207 071

CS 206 585

AUTHOR Roos, Michael E.
TITLE Syntactic Maturity and Grading: A Correlational Study.
PUB DATE Jul 81
NOTE 17p.; Paper presented at the Annual Meeting of the Wyoming Conference on Freshman and Sophomore English (10th, Laramie, WY, July 6-10, 1981).
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Correlation; Essays; Evaluation Methods; *Grading; Higher Education; Literary Styles; Sentence Combining; *Sentence Structure; *Structural Analysis (Linguistics); *Syntax; Two Year Colleges; Writing (Composition); *Writing Evaluation; *Writing Research
IDENTIFIERS Clauses; *Syntactic Maturity; T Units

ABSTRACT

A study was conducted to determine whether a significant correlation existed between an essay's letter grade and five important factors of syntactic maturity (clause length, t-unit length, sentence length, clauses per t-unit, and t-units per sentence). A total of 45 essays--15 A, 15 B, and 15 C papers as graded by five junior college English teachers--were assigned a point value and analyzed for a correlation between the grade and each of the five syntactic maturity factors. The results indicated a slight correlation between the grades and the number of clauses per t-unit and the number of t-units per sentence, suggesting that the teachers were influenced somewhat in their grading by the writer's ability to build compound sentences. The second portion of the study was an analysis of the difference between each of the three grade groups. Here, the number of words per sentence was found to be the most significant factor. The significant difference between the A and the C papers was possibly due to the fact that the A students connected independent clauses with semicolons more frequently than did the other students and thus produced much longer sentences and more t-units per sentence. While the results infer that sentence combining exercises may help increase the quality of a student's writing, factors of syntactic maturity overall had only a low correlation with the quality of writing. (HTH)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED207071

Michael E. Roos
 Assistant Professor of English
 Raymond Walters College
 University of Cincinnati
 Cincinnati, Ohio 45236
 Phone: 745-4264 (office)
 731-8455 (home)

"PERMISSION TO REPRODUCE THIS
 MATERIAL HAS BEEN GRANTED BY

Michael E. Roos

TO THE EDUCATIONAL RESOURCES
 INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
 NATIONAL INSTITUTE OF EDUCATION
 EDUCATIONAL RESOURCES INFORMATION
 CENTER (ERIC)

- This document has been reproduced as received from the person or organization originating it.
- Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Syntactic Maturity and Grading: A Correlational Study

While research in the teaching of English composition has been resistant to the use of empirical studies, the past fifteen years have seen significant empirical studies that imply that good writing can be measured quantitatively. The two most significant measurements have been found to be the student's average words per t-unit and the student's average words per clause. Simply put, the studies indicate that the greater these averages are, the more mature the student is as a writer. The outgrowth of these studies, of course, has been sentence combining. Evidence from studies of sentence combining has shown quite conclusively that it is indeed a very efficient means of increasing a student's average clause and t-unit length. But the studies have not gone very far in showing a relationship between these quantitative measurements and the overall quality of a student's writing, based on its content, organization, style, and mechanics. The fact is that unless such a relationship can be demonstrated, the case for a composition course based solely on sentence combining is weakened.

Consequently, I designed a study that would determine if a significant correlation exists between an essay's grade and five important factors of syntactic maturity. I began with the hypothesis that the correlation between them would be high and significant. I ended with the realization that the mathematical correlation between the grade and any one of these factors is low, but there is, nonetheless, a statistically significant difference between good writers and average writers at the college level. My findings are not inconsistent with

CS 206585

previous research in this area and in fact reinforce it.

The first really significant quantitative research in English was a study by Kellogg Hunt, "a search for developmental trends in the frequency of various grammatical structures written by students of average IQ in the fourth, eighth, and twelfth grades."¹ In addition, Hunt also analyzed samples of writing from Harper's and Atlantic magazines, in order to establish frequency counts for the most mature types of writing. His results are summarized in the table below:²

	words/ clause	words/ t-unit	words/ sentence	clauses/ t-unit	t-units/ sentence
Grade 4	6.6	8.6	13.5	1.30	1.60
Grade 8	8.1	11.5	15.9	1.42	1.37
Grade 12	8.6	14.4	16.9	1.68	1.17
Superior Adults	11.5	20.3	24.7	1.78	1.23

Hunt's study indicated that t-unit length is the most significant measure of maturity until the twelfth grade, when clause length becomes the most significant. The wide gap Hunt found between twelfth graders and superior adults in clause length left open the opportunity for studies to show some means of narrowing the gap, implying that if a student could increase his average words per clause by three words he would be in a class with professional writers. As a direct consequence, research in sentence combining was begun to find the magic bridge over the gap.

John Mellon was the first to test the effectiveness of sentence combining exercises in increasing syntactic maturity, studying Boston seventh grade students. Mellon's findings showed not surprisingly that students who received instruction

in sentence combining improved in average length of t-unit much more than students who received instruction in traditional grammar.³ Though his findings were significant, a weakness of Mellon's study was that it did not distinguish between sentence combining exercises and other forms of actual writing practice in similar quantities. In other words, the gains in the experimental group might be explained simply by the fact that they got much more practice in actual writing than the control group. In addition the study made no effort to determine if the overall quality of a student's writing was improved more by sentence combining than by other techniques of teaching English.

Frank O'Hare recognized the latter weakness and included in his study a qualitative judgment of essays in the experimental and the control group.⁴ His subjects were seventh graders at the Florida State University High School. O'Hare deliberately structured his study so that it would be comparable with Mellon's, but he made some significant changes, in addition to the qualitative judgment, which tended to rectify some of the weaknesses of Mellon's study. Again, significant improvement in syntactic maturity was apparently produced by the sentence combining exercises. And, in fact, the experimental group seemed to have been raised to the twelfth grade level or beyond, using Hunt's data as normative. The next step was to assess the overall quality of the writing of the two groups. Given the probability that, if there were no differences between the groups of essays, experimental group essays would be chosen exactly half the time, O'Hare found that, in fact, essays of the experimental group were selected as better 169 times compared to an expected 120. O'Hare's study was clearly the most significant one to date in determining the real value of sentence combining. However, there was no indication of the bias of the eight teacher/evaluators of the writing samples and to what extent they were aware of the nature of the study being done. For obvious reasons, it is essential that such evaluators be

totally objective, completely unaware of the nature of the study. In addition, O'Hare did not do a pre-test qualitative analysis of the two groups. Thus, we don't know if the experimental group were already better overall writers when they came into the course. Nevertheless, O'Hare's study has strong implications that sentence combining deserves a role at least in the seventh grade English curriculum.

Donald Daiker, Andrew Kerek, and Max Morenberg published a report in 1978 of research that was designed to show the same effects at the college level that Mellon's and O'Hare's studies showed at the seventh grade level.⁵ Daiker, Kerek, and Morenberg studied twelve sections of the basic composition course at Miami University, six in the experimental group, six in the control group. Their results are shown in the table below:⁶

Pre-test Scores				
	Experimental	Control	Difference	T-value
words/clause	8.75	8.30	-.05	-.32 (NS)
words/t-unit	15.31	14.99	.32	1.00 (NS)
clauses/t-unit	1.76	1.72	.04	1.43(NS)
Post-test Scores				
	Experimental	Control	Difference	T-value
words/clause	9.64	8.67	.97	5.58 (p<.001)
words/t-unit	16.05	14.95	1.10	3.37 (p<.001)
clauses/t-unit	1.69	1.73	-.04	-1.45 (NS)

Though their study did not show anything like the phenomenal growth rate in their students that O'Hare's study did, clearly they have shown that a sentence combining curriculum increases the syntactic maturity of college Freshmen.

They tell us in a footnote that a qualitative comparison between the two groups was done, though they published no details of that portion of their study. They assert that the experimental group was found to write free compositions judged by experienced English teachers to be significantly superior to these of the control group.⁷ But again, as with O'Hare's study, no indication of a pre-test qualitative comparison was given. Without such a test, we cannot say with assurance that sentence combining exercises raise the overall effectiveness of a writer.

My study was designed to inject the element of an essay's overall quality directly into the analysis of the factors of syntactic maturity. I decided that the primary means of doing this would be to perform an analysis of the correlation between the grade received by an essay and Hunt's five factors of syntactical maturity: clause length, t-unit length, sentence length, clauses per t-unit, and t-units per sentence. In addition, I would look at the averages for these factors in all the papers that received a certain grade, testing the differences between each grade level for statistical significance. The third portion of my study was a survey of the teachers involved, to determine the extent to which they feel they are influenced by a writer's style, i.e., his syntactic maturity. The sum of these three portions of the study, it was hoped, would provide a clear picture of the relationship between a writer's syntactic maturity and his overall effectiveness as a writer.

I wanted the sample to include essays assigned and graded by experienced teachers, whose sound judgment in evaluating essays had been demonstrated. Because of my location at Raymond Walters College, a small two-year school, I was limited to English faculty members here. I was, however, able to choose five able and respected teachers, four of whom are tenured, the fifth being a promising young teacher in her third year on the faculty here. Three of the

teachers were male, two female. They ranged in age from 27 to 45. My survey revealed that none of the teachers used sentence combining extensively in class, though **two** of them had dabbled with its use, without indulging in a formal sentence combining text for their students. Because of the small size of the department here, it was impossible to get five teachers of the same course. Instead, I got a representative sampling of the three kinds of English courses taught at Raymond Walters: lower track Freshmen, upper track Freshmen, and Sophomore literature students. Two of the five teachers were from the lower track English courses, which consist of students scoring below 350 (verbal) on the Scholastic Aptitude Test, although some higher scoring students choose to take this course for a variety of reasons. In addition, two of the five were from the upper track English course. The fifth teacher was from a Sophomore literature course in the twentieth century novel. Thus, I feel the sample of teachers is representative of English teachers at large and should reflect accurately grading practices throughout the profession.

What those real grading practices are, however, has too often been a mystery, so I directed a section of my teacher survey precisely to that problem. Each teacher was asked to estimate roughly what percentage of an essay grade is determined by a student's proficiency in each of four areas: content, organization, style, and mechanics. The results are indicated in the table below:

Teacher	A	B	C	D	E	Average
content	30	25	60	30	20	35
organization	20	50	15	20	20	25
style	10	0	10	10	0	6
mechanics	40	25	15	40	50	34

It should be noted that Teacher C taught the Sophomore literature course, which naturally is more content oriented than a Freshman course, a fact reflected in the 60 percent weight he placed on content in grading. The other teachers were fairly consistent. The area of most interest to me, however, for the purposes of this study, was style, which should involve all the factors of syntactic maturity. Interestingly, no teacher rated it higher than 10 percent, and two of the three teachers insisted it had no effect on a paper's grade. One of them commented on the survey, "I don't think we have students who are ready to work on style." Consciously, then, these five teachers placed an average of only six percent of an essay grade under the category of style.

Each of the five teachers was asked to submit a total of nine graded essays, randomly selecting three A's, three B's, and three C's from the same assignment. All of the essays were written in class with no opportunity for extensive revision, though two of the sets were from final exams, which allowed the students two hours to finish instead of the normal one hour. The study was limited to the top three grades because it was assumed that any significant results for these grades would hold true as well for D's and F's. This assumption may not be true, however, and future studies should explore these grades also. The total of 45 essays analyzed thus included 15 A's, 15 B's, and 15 C's, selected at random by different teachers from a variety of English students at Raymond Walters College.

Once the essays were collected from the teachers, I scored each one according to the five factors of syntactic maturity previously mentioned. I then assigned a point value to the grade the paper had received: 12 points for an A+, 11 for an A, 10 for an A-, and so on, down to 4 for a C-. This part of the research then involved an analysis of the correlation between the grade and each of the five factors.

The results are summarized below. Keep in mind that a correlation coefficient indicates the extent to which we can predict one variable if we know the other.

Preliminary Correlational Results

	Correlation Coefficient	T-value
words/clause	.25	1.69 (p<.05)
clauses/t-unit	.24	1.62 (p<.05)
words/t-unit	.40	2.86 (p<.005)
t-units/sentence	.35	2.46 (p<.01)
words/sentence	.55	4.29 (p<.0005)

Also recall that Hunt had concluded that, for writers beyond the twelfth grade level, words per clause is the most significant measure of syntactic maturity, followed in order by words per t-unit and clauses per t-unit. Words per sentence and t-units per sentence were determined to be least significant of the five. My preliminary results showed an opposite effect, however. While words per clause showed a small, insignificant correlation with the grade, words per t-unit showed a higher significant correlation, and words per sentence an even higher, quite significant correlation. This led to an absurd conclusion that a writer need only build long sentences to get a high grade on an essay.

But these preliminary results were very misleading and required a much closer analysis to determine the true correlations involved. First I had to recognize that words per t-unit and words per sentence are dependent on the other measurements, expressed mathematically $(\text{words/clause}) \times (\text{clauses/t-unit}) = \text{words/t-unit}$ and $(\text{words/t-unit}) \times (\text{t-units/sentence}) = \text{words/sentence}$. Consequently, the effects of words per clause and clauses per t-unit are included in the total effect of words per t-unit, and all the factors effect words per

sentence. The true correlation, thus, must be found by using the formula for partial correlations for each factor, that is, with all the other factors held constant. Only then can we find the real correlation of each factor without the additive effect of the others included. The results of the partial correlations are summarized in the table below. What we find should be a more accurate picture of the real effect any one of these factors has on the grade, though there are still some puzzling aspects. There is a decline, as would be expected, in the true correlation from words per clause to words per t-unit to words per sentence, with words per clause now the most statistically significant.

	True correlation
words/clause	.25 ($p < .05$)
clauses/t-unit	.24 ($p < .05$)
words/t-unit	.20 (not significant)
t-units/sentence	.35 ($p < .01$)
words/sentence	.14 (not significant)

The puzzling results, however, come under clauses per t-unit and t-units per sentence, both of which have an unexpectedly high true correlation. Hunt said that clauses per t-unit is less significant than words per t-unit as a measure of maturity. Here it seems to be more significant. The most surprising of all is t-units per sentence, which is a measure of the extent to which a writer uses compound sentences. Common sense tells us, and Hunt's study showed conclusively, that very immature writers can have high ratios of t-units per sentence (sentences such as: I went to the store, and I bought some eggs, and then I came back home, and Mom made me wash for supper). But Hunt's study did show that despite the fact that the ratio of t-units per sentence declines from grades four through

twelve, it does climb back up somewhat, though not significantly, for superior adults. At any rate, it appears that these five English teachers were influenced in their grading by a writer's ability to build compound sentences. I would tentatively interject here that as I was scoring these essays it appeared that the better students used semicolons to separate independent clauses (and thus make compound sentences) much more frequently than other students. This may reflect the emphasis in Freshman English courses on using semicolons to correct comma splices.

The overall conclusion from this portion of the study, however, has to be that none of these factors plays a really significant role in determining the grade an essay receives. The correlations remain fairly low, and when we do an analysis of variance, we arrive at only 30 percent of the variation in the grade explained by these factors. That leaves 70 percent unexplained to be determined by other factors, including, it is presumed, content, organization, and mechanics, which are not addressed directly in a sentence combining curriculum. More correlational studies really need to be done, however, so that we can compare factors of syntactic maturity with other individual factors to determine the relative importance of each. We do not know, for example, if organization or the number of mechanical errors would show more correlation with the grade than syntactic maturity appears to.

The final portion of my study was devoted to finding the averages of all the papers in each grade group. Here plus and minus were disregarded in the grades, so that I had three grade groups: A's, B's, and C's. The results are tabulated below. T-tests for statistical significance were applied to the differences between grade levels. Between A's and B's, the difference of .59 words per clause was not found to be significant for a sample of this size. Likewise, the difference of .07 clauses per t-unit was not found to be

significant. However, the differences in the other three factors were all found to be significant at or beyond the .025 level, with the most significant factor

Total Averages by Grade Level					
	words/ clause	clauses/ t-unit	words/ t-unit	t-units/ sentence	words/ sentence
A papers	9.59	1.65	15.84	1.26	20.02
B papers	9.00	1.57	14.18	1.11	15.71
C papers	8.63	1.60	13.85	1.14	15.82

being words per sentence. The results are summarized below:

Difference Between A's and B's				
	A papers	B papers	Difference	T-value
words/clause	9.59	9.00	0.59	1.53 ($p < .10$)
clauses/t-unit	1.65	1.57	0.08	1.19 (not significant)
words/t-unit	15.84	14.18	1.66	2.30 ($p < .025$)
t-units/sentence	1.26	1.11	0.15	3.66 ($p < .005$)
words/sentence	20.02	15.71	4.31	4.90 ($p < .0005$)

An analysis of the difference between B's and C's revealed somewhat different results. None of the differences were found to be significant: in fact, C's were found to score slightly, though insignificantly, higher than B's in words per sentence, clauses per t-unit, and t-units per sentence. These facts especially weaken the notion that words per sentence and t-units per sentence are important determinants of a paper's grade, though they may indeed be important for an A paper. The fact is there is no appreciable difference between B's and C's on these three factors and only a slightly higher

difference in words per t-unit. Here at last we find words per clause to contain the most appreciable difference between the two grades, though still not at a statistically significant level. The results are shown below:

Difference Between B's and C's				
	B papers	C papers	Difference	T-value
words/clause	9.00	8.63	.37	1.11 (not significant)
clauses/t-unit	1.57	1.60	-.03	(negative; not significant)
words/t-unit	14.18	13.85	.33	.73 (not significant)
t-units/sentence	1.11	1.14	-.03	(negative; not significant)
words/sentence	15.71	15.82	-.11	(negative; not significant)

Only when we compare A's and C's do we find statistically significant differences in all factors except clauses per t-unit. The results are indicated below:

Difference Between A's and C's				
	A papers	C papers	Difference	T-value
words/clause	9.59	8.63	.96	7.80 ($p < .0005$)
clauses/t-unit	1.65	1.60	.05	.82 (not significant)
words/t-unit	15.84	13.85	1.99	2.69 ($p < .025$)
t-units/sentence	1.26	1.14	.12	3.08 ($p < .005$)
words/sentence	20.02	15.82	4.20	4.50 ($p < .0005$)

Just as Hunt predicted, words per clause is shown to be the most significant of all the measurements, though for this study we are still left with surprisingly high significance for words per sentence and t-units per sentence, at least when comparing A's with all other papers. Again, my tentative explanation would be that the A students tended to connect independent clauses with semicolons much more frequently than other students, resulting in much longer sentences and more t-units per sentence.

At this point it is interesting to compare my findings to Hunt's, which have been viewed as normative by virtually all researchers in the field, and to Daiker, Kerek, and Morenberg's, which dealt with students at the same level as those in my study (with the exception of the Sophomores). My group of C paper's, according to Hunt's data, are found to be at or slightly below the twelfth grade level in all the factors.³ The B's are not much different, except that they are a little above the twelfth grade level in words per clause. The A's, however, are well above the twelfth grade level, as they should be, in all aspects except clauses per t-unit, which has not yielded very significant results for us in any sense. The A's are still well below superior adults in all respects except t-units per sentence, where they actually exceed superior adults slightly. Again, the emphasis on semicolons might be responsible. In a comparison of my findings with those of the Miami study, the C's are found to be slightly lower than the Miami students when they were pre-tested in the three factors studied there. However, the A papers' averages are virtually equal to the post-test scores of the Miami experimental group, those students who received instruction in sentence combining. Their control students remained on the post-test comparable to our B and C students. This is perhaps one of the most significant findings to come out of my study, and it may lend some support in the area of qualitative analysis to the Miami study. The inference, of course, is that sentence

combining exercises may be able to raise the student's overall quality of writing significantly, perhaps even as much as from C quality to A quality. Such a conclusion would be significant indeed, but I hasten to add that the possible effects of different student populations and writing assignments may account for differences or apparent similarities between my results and theirs. Nonetheless, there is nothing in my study to indicate that sentence combining would not be helpful to college English students.

My study has not been a perfect one; ideally, such a project would involve more teachers, perhaps ten, and would include analysis of D's and F's along with the other grades. However, my study has shown some significant results and, at the very least, has been promising enough to indicate the need for further research of this type. It has shown that factors of syntactic maturity have only a very low mathematical correlation between them and a judgment of the overall quality of a piece of writing. Predicting a writer's grade from factors of syntactic maturity alone would not yield a very high degree of success. For example, from my study, should we demote the A- student who had only 7.41 words per clause to a C or below? By the same token, should we raise the C student, of the same teacher, who had 9.85 words per clause to an A? Obviously, the teacher of the two students doesn't think so, and I don't either, judging from my own experience in grading essays. Syntactic maturity is only one factor in a writer's effectiveness, and an ability to build long clauses and t-units alone will not make a writer proficient. What we do not know yet is the precise correlation other factors have with an essay's grade and whether or not they are more significant than syntactic maturity. It may well be that the evaluation of writing is such a complex process that no one of the four factors plays a strong role in determining the grade. It does appear at this point, however, that syntactic maturity should not be ignored as one of those factors and that English depart-

ments need to consider including sentence combining as a part of a comprehensive college Freshman English curriculum, which should include instruction in content development, organization, mechanics, and style, all of which seem to play some role in determining how well a student writes.

Notes

¹Kellogg W. Hunt, Grammatical Structures Written at Three Grade Levels (Champaign, Ill.: National Council of Teachers of English, 1965), p. 1.

²Hunt, p. 56.

³John C. Mellon, Transformational Sentence Combining (Champaign, Ill.: National Council of Teachers of English, 1969), p. 31.

⁴Frank O'Hare, Sentence Combining: Improving Student Writing without Formal Grammar Instruction (Champaign, Ill.: National Council of Teachers of English, 1973), p. 35.

⁵Donald Daiker, Andrew Kerek, and Max Morenberg, "Sentence Combining and Syntactic Maturity in Freshman English," College Composition and Communication, 29(February, 1978), 36-41.

⁶Daiker, pp. 39-40.

⁷Daiker, footnote p. 38.

⁸Keep in mind that Hunt's norms are from 1965, and these, like SAT scores may have declined since then.